Estimate of model error covariance Q for weak-constraint 4D-Var

Currently the operational implementation of 4D-Var at ECMWF uses strong-constraint 4D-Var [4][5][7]. Strong-constraint 4D-Var relies on the assumption that the numerical model's representation of the evolution of atmospheric flow is perfect, or at least that the model errors are small enough to be neglected compared to other errors in the system [3]. Errors in observations and background state are accounted for using the **R** observation and the **B** background error covariance matrices. As other aspects of data assimilation processes have advanced, the validity of this perfect model assumption becomes more questionable and limits the length of the analysis window to roughly 12 hours. Weak-constraint 4D-Var relaxes the perfect model assumption by explicitly representing model error as part of the 4D-Var control variable. The model is now only a weak constraint on the system. However, a model error covariance matrix is required. Here, a new model error covariance matrix based on statistics from parametrised model error schemes is proposed for use in the short forecast.

Model Error Formulation

Model error contains both random and systematic (or even constant) components. To simplify the problem we consider the model error to be constant by intervals. For the work presented here, the interval we have chosen is one constant forcing for the whole 12 hour assimilation window, at the other extreme we could (in principle: it is not yet technically possible in the IFS) have chosen to have the interval as short as a model time step; this would be the full 4D problem.
The 4d-Var cost function we are considering is:

$$J(\mathbf{x}_0, \eta) = \frac{1}{2}(x_0 - x_b)^{\mathrm{T}}\mathbf{B}^{-1}(x_0 - x_b) + \frac{1}{2}\sum_{k=0}^{N}(\mathcal{H}_k(x_k) - y_k)^{\mathrm{T}}\mathbf{R}_k^{-1}(\mathcal{H}_k(x_k) - y_k) + \frac{1}{2}(\eta - \eta_b)^{\mathrm{T}}\mathbf{Q}^{-1}(\eta - \eta_b) \tag{1}$$

where η_b is the mean model error, x_k is the state at time k with $x_k = \mathcal{M}(x_{k-1}) + \eta$ representing the state at time k resulting from the forced model integration from time $t = 0$ to $t = k$, η represents the instantaneous model error. Observations and model error are assumed uncorrelated in time.[6]

Calculation of model error covariance matrix (Q)

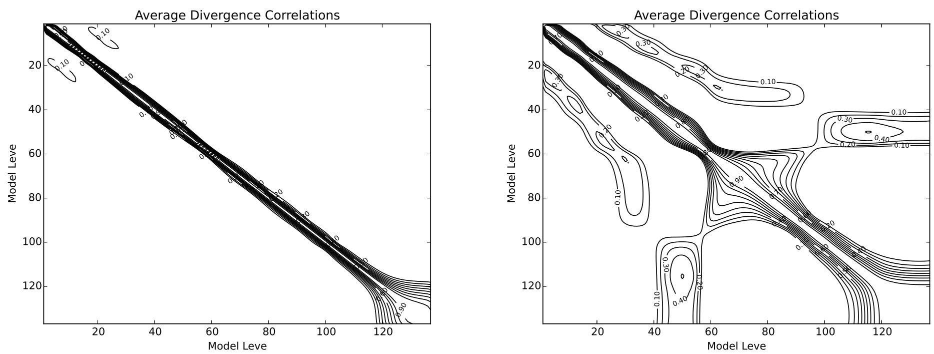
In order to calculate the new Q matrix, statistics are generated from special runs of the Ensemble Prediction System (EPS) in which initial perturbations are removed. In these runs, members diverge from each other due to their different realisations of parametrised model error (SKEB and SPPT). The differences between members after 12 hours of model integration give an estimate of the integrated effect of model error over 12 hours; from which statistics appropriate for use in 4D-Var can be calculated. These statistics are used to construct a covariance model similar to that described by Derber and Bouttier [1] for background error covariances. (Note, however, that the model error covariance matrix we have constructed does not include a balance operator: model errors for different variables are assumed to be uncorrelated.) This method of generating model error covariance statistics provides greater consistency between the approaches taken to representing model error in the 4D-Var and EPS systems than previous methods.

EPS experiment description:

- 50 member ensemble + control;
- $\mathrm{T_L}$ 399 resolution;
- 12 hour forecast;
- cycle 40R3;
- 20 days of forecasts (2013083100 - 2013091900);
- identical initial conditions (ensemble members are not perturbed).

Initial experimentation using the new $\mathbf{Q}$ matrix suggested that the implied variances of model error were too large. Weak-constraint analyses using the matrix were found to have very small initial increments as 4d-Var found it less costly to nudge the state towards the solution via a model-error correction than to correct the initial state. In order to select a reasonable magnitude of the $\mathbf{Q}$ matrix we looked at the minimisation of the cost function for a range of different values of multiplicative factors between 0 and 1 and choose a value for which the model error $\mathbf{Q}$ term has an influence but does not dominate over the other error terms.

For the multiplicative factor value chosen ($\alpha = 0.2$), a 4D-Var weak-constraint assimilation experiment was run. The 4D-Var model error estimates η from this experiment were then used to calculate a covariance matrix that could be compared with the $\mathbf{Q}$ matrix used in the assimilation. The experiment was run for 90 days with 12-hour assimilation windows starting at 0900 and 2100; both times were used for calculating the model error covariance estimates.



(a) 12 hour stochastic

(b) Weak constraint

Figure 1: Divergence model error average vertical correlations for 12 hour stochastic model error and weak-constraint 4D-Var model error estimate. Contour interval is 0.1 for both figures.

In fig. 1 we see the average vertical correlations of estimated model error for divergence. The 4D-Var weak-constraint model error covariance estimate (right panel) does not retain the same structure as the stochastic model error covariance, $\mathbf{Q}$ (left panel). In particular, we see some unexpected correlations between levels that are far apart.
By looking at the geographical location of these correlations we saw a clear pattern over North America and Europe corresponding to areas with a high number of aircraft observations. This suggests that 4D-Var is misinterpreting aircraft observation error or bias as model error. In order to avoid this interaction with aircraft observations, subsequent experiments restricted the effect of model error to be active only above 100hPa.
A CY41R2 $T_{co}1279$ experiment was run. Forecast skill scores were verified against own analysis and also against GPS Radio Occultation (GPSRO) observations. The verification against own analysis in the northern hemisphere (fig. 2) showed a significant reduction in RMS error at 100hPa. GPSRO verification in the stratosphere showed a change in the bias structure throughout the forecast. In the northern hemisphere bias is slightly improved at all levels, in the tropics it is largely unchanged and in the southern hemisphere the results are mixed (but the differences from the control in this region are very small).

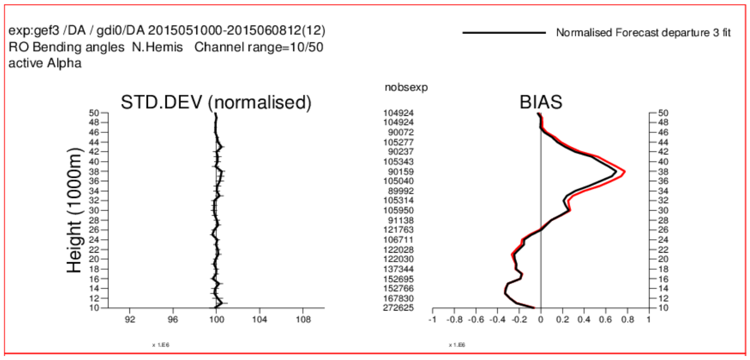


Figure 2: GPSRO verification northern hemisphere

We hope to introduce this configuration of weak-constraint 4D-Var with model error forcing above 100hPa into CY43R1. Understanding and reducing aliasing of observation error is critical to plans to extend the model error representation to all model levels, and will require improvements to the representation of systematic observation error (in particular biases in aircraft data). Finally, we plan to extend our research to encompass the random component of model error once the technical facility to represent it in 4d-Var exists. For this, we rely on ongoing developments within the OOPS project.

Bibliography

[1] Derber J., Bouttier F., *A reformulation of the background error covariance in the ECMWF global data assimilation system*, Tellus, 51A, 195-221, 1999.

[2] Elías Valur Hólm, Yannick Trémolet, *Investigation leading to turning off model error for L137 in CY38R2*, Research Department Memorandum, RD 13-422, October 29, 2013.

[3] Yannick Trémolet, *Model error estimation in 4D-Var*, Research Department Memorandum, RD 07-520, March, 2007.

[4] Rabier, F., Jrvinen, H., Klinker, E., Mahfouf, J.-F. and Simmons, A. (2000), *The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics.* Q.J.R. Meteorol. Soc., 126: 11431170., doi: 10.1002/qj.49712656415.

[5] Mahfouf, J.-F. and Rabier, F. (2000), *The ECMWF operational implementation of four-dimensional variational assimilation. II: Experimental results with improved physics.* Q.J.R. Meteorol. Soc., 126: 11711190., doi: 10.1002/qj.49712656416.

[6] Mike Fisher, Yannick Trémolet, Harri Auvinen, David Tan and Paul Poli, *Weak-Constraint and Long-Window 4D-Var.* Research Department Memorandum, RD-655, December 2011.

[7] Klinker E, Rabier F, Kelly G, Mahfouf JF. (2000) *The ECMWF operational implementation of fourdimensional variational assimilation. III: Experimental results and diagnostics with operational configuration.* Q.J.R. Meteorol. Soc., Apr 1;126(564):1191-215.